DATS 6101 Introduction to Data Science

Project 2 Outline (Spring 2020)

Goal: Build and interpret models that answer your SMART research questions from data.

I. Development of a **research driven question (SMART)** focused on a dataset either inside of R or one of your choosing from any online sources (3000+ observations). It can be the same dataset as your project 1. The questions would be different from project 1 however.

II. Provide an **R-markdown file**, knitted into **HTML**, which shows the R-code and brief explanations for the technical work in your project. (**Also** submit your **data file**, or give the online source url.) It should include:
    - Summary of the dataset (just basic summary(dataframe) will do here)
    - Model(s) used
    - Model evaluation(s) and comparison

III. Write a roughly 10-page (definitely no more than 4000 words) summary of the research and EDA process of your project. The summary should be prepared in **R-markdown**, and knitted into **HTML**. You may take some of the work in part II (such as graphs and results) to include here. They can overlap. This summary is to-be presented to your boss, your client, or to-be submitted for publication in journals. Potential area of topics to address in this summary may include:
    - Some basic EDA.
    - How did you select and determine the correct model to answer your question?
    - What predictions can you make with your model? Examples.
    - How reliable are your results?
    - What additional information or analysis might improve your model results or work to control limitations?
    - References (APA style preferred)

IV. Develop a **15-20 minute** presentation for the team that effectively communicates the results of these models and interpretation to the audience.

Grading:

I. *
II. 40%
III. Together with part I, total of 40%
IV. 20% (Individually graded)

Grades for parts I through III are team-based. But I reserve the rights to award different grades to team members if there is evidence of unfair contribution proportion within the team.